

**APPENDIX A**  
National TAC Statement on WASL Quality

January 26, 2004

TO: Dr. Terry Bergeson  
State Superintendent of Public Instruction

FROM: National Technical Advisory Committee

RE: Review of the Quality of the Washington Assessment of Student Learning

The National Technical Advisory Committee (TAC) used the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], National Council on Measurement in Education [NCME], 1999) as a basis for making professional judgments about the reliability and validity of scores from the *Washington Assessment of Student Learning* (WASL).

The TAC has reviewed the technical reports for years 1999 through 2002, annual reports on linking and equating and related research. These reports provide evidence regarding the validity of inferences made from scores about the level of student achievement on the *Essential Academic Learning Requirements* (EALRs). Reports included, but were not limited to, information on:

- Test development methods.
- Item content review.
- Item analyses.
- Methods for setting performance standards.
- Evidence for score and inter-rater reliability.
- Evidence for validity of scores.
- Methods of scaling and equating.
- Annual descriptive data regarding statewide performance on the WASL tests.
- Statewide performance for students in categorical programs, by ethnic groups, and by gender.

Based on these documents and other studies and reports presented to the National (TAC), we can make the following statements:

- This committee concludes that the WASL meets the relevant standards of validity as prescribed by the national *Standards for Educational and Psychological Testing* (AERA, APA, and NCME 1999).
- The test design and the test and item specifications for each individual year indicate that the items on the test adequately represent the *EALRs* for the state of Washington.
- The level of validity and reliability for reporting individual student and school results is acceptable for reading, mathematics, and writing.
- Methods used to develop and improve the quality of items and the tests are consistent with standard technical practices for development of criterion-referenced tests.
- The methods of item development, analysis, and selection are consistent with standard practices for the construction of criterion-referenced tests.
- The item development and review processes have contributed significantly to the content validity of the assessments.
- The bias sensitivity review and Differential Item Functioning (DIF) analyses that were performed ensure that adequate attention was given to issues of fairness across subgroups.
- The state has used procedures for setting performance standards (including determining method, selecting panelists, summary information) that meet or exceed expectations established by the *Standards for Educational and Psychological Testing* (AERA, APA, and NCME).
- The TAC has approved the procedures to be used to review standards for Reading, Mathematics, and Writing during the spring of 2004.
- Scaling and equating procedures reflect standard practices in measurement. These procedures have been annually reviewed and approved by the TAC.

- The WASL reading and mathematics scales appear to be quite stable over time based on equating studies and other research presented to this committee.
- Equating procedures ensure that the performance standards have remained stable within subject, within grade over time.
- The WASL writing scoring procedures provide stability to the writing scale.
- Inter-rater reliability data suggest that scoring methods are carefully controlled such that there is a high level of inter-rater agreement and that students' total scores are likely to be about the same regardless of the raters.
- Given the opportunities for multiple retakes, the Grade 10 WASL scores are sufficiently reliable and valid to award the Certificate of Mastery.

The observations and recommendations offered by the TAC in this document take into account the procedures and strategies implemented by the Office of Superintendent of Public Instruction (OSPI) and their contractors over the life of the WASL program and intended for the future. OSPI has consistently demonstrated willingness to act on recommendations of the TAC designed to ensure a high quality program. The WASL assessment program meets or exceeds technical requirements outlined by *Standards for Educational and Psychological Testing* (AERA, APA, and NCME 1999).

*Recommendations for future action:*

Given that the award of the Certificate of Mastery is a high stakes decision for individual students, the National TAC has discussed and recommends incorporating the following components:

- A partially compensatory decision model or a variation that allows for blending of scores.
- Alternative options for students from special populations.
- An appeal process.
- Policy and support for instruction and remediation.

The TAC encourages OSPI to continue to support educational policies and programs that provide all students with an adequate opportunity to learn the EALRs during their K-12 experience. This expectation is consistent with professional standards of fairness and validity for the implementation of a high- stakes assessment system.

Attachment

## **National Technical Advisory Committee Members**

### **Patricia Almond, Oregon Department of Education and University of Oregon**

Dr. Almond was director of alternate assessments and testing accommodations for the Oregon Department of Education between 1996 and 2004. She directed the design, development, and implementation of Oregon's alternate standards and assessments for students with significant disabilities, including: career and life role assessment systems, extended reading, extended writing, and extended mathematics assessments. Among her many publications, Dr. Almond wrote the chapter on participation and inclusion of all students including students with disabilities in large scale assessment for the book *Large Scale Assessment for All Students*, G. Tindall and T. Haladyna (Eds.) published by Lawrence Erlbaum Associates. She chaired the Research Study Group the States Collaborative on Assessing Student Standards: Assessing Special Education Students between 1997 and 2003. She serves as an affiliate with National Center on Educational Outcomes and Co-Directs Project CASIAS: Comprehensive Assessment Systems: Including All Students, a funded to the University of Oregon by the US Office of Special Education Programs. Currently, she is the principal researcher for Colorado's Enhanced Assessment Grant, investigating the technical adequacy of performance assessments for students with significant cognitive disabilities, a federally funded project under the No Child Left Behind act.

### **Peter Behuniak, University of Connecticut**

Dr. Behuniak is on the faculty of Educational Psychology: Measurement, Evaluation, and Assessment; formerly Director of Assessment for the Connecticut Department of Education. He has written and presented extensively on issues related to large-scale educational assessment. He has served as the chairperson of the Council of Chief State School Officers Consortium on Technical Issues in Large-Scale Assessment and Studies Panel and the South Carolina technical advisory committee.

### **Richard Duran, Professor of Education, Gevirtz Graduate School of Education, University of California, Santa Barbara**

Professor Duran previously worked at the Educational Testing Service in Princeton where he conducted investigations and published research findings on the validity of the SAT, GRE, and TOEFL tests. Since joining the UCSB GGSE faculty in 1984, he has conducted research investigating learning and culture (funded by Center for Research on Education Standards and Testing, and the Center for Research on Education of Students Placed At-Risk), investigations of children's learning in after school computer club settings (supported by UC LINKS), new models of literacy achievement arising through children's computer club participation (funded through the Mellon Foundation), and studies on how immigrant parents can help their children conduct research and publish projects (funded by the Center for Research on Education, Diversity, and Excellence). In addition to service on the Washington State Assessment Technical Advisory Committee, Prof. Duran serves on similar committees for New York and Texas, and on the California English Language Development Test Technical Advisory Committee.

### **Professor George Engelhard, Jr. Ph.D., Emory University**

Dr. George Engelhard, Jr. is a professor of educational measurement and policy in the Division of Educational Studies at Emory University (1985 to present). Professor Engelhard received his Ph.D. (1985) from The University of Chicago in the MESA (measurement evaluation and statistical analysis) program with a concentration in the sociology of education. His research has focused on a variety of topics including: the role of assessment in improving educational processes; Rasch measurement theory; rater-mediated assessments; history of measurement theory; assessment of written composition; development of affective human characteristics; teacher assessment; and certification testing. Among his current projects is research on the following: models of judgment and educational assessment; examination of rater effects in a variety of contexts; and standard setting on high-stakes performance assessments. Professor Engelhard is co-editor of four books, and the author or co-author of over 100 journal articles, book chapters, and monographs. He serves on several national technical advisory committees on educational measurement policy including Georgia, Michigan, New Jersey, Ohio, Oregon, and Washington. He has served as President of the Georgia Educational Research Association (1992-93). He has received several awards and fellowships including a National Academy of Education/Spencer Foundation Fellowship Award, a Lilly Post-Doctoral Teaching Award, and a Writing Across the Curriculum Project Award. [20 page curriculum vitae is available upon request]

### **Robert Linn, University of Colorado**

Robert Linn is distinguished professor of education in the research and evaluation methods program. Professor Linn is a member of the National Academy of Education and a lifetime National Associate of the National Academies. He has been an active member of the American Educational Research Association (AERA) for more than 30 years and has served as vice president of the AERA Division of Measurement and Research Methodology and vice chair of the joint committee that developed the 1985 Standards for Educational and Psychological Testing. He is the immediate past president of AERA. He is a past president of the National Council on Measurement in Education (NCME), past editor of the *Journal of Educational Measurement*, and editor of the third edition of *Educational Measurement*, a handbook sponsored by NCME and the American Council on Education. He was chair of the National Research Council's (NRC) Board on Testing and Assessment and currently serves on the NRC's Board of the Center for Education. Dr. Linn is also the co-director of the National Center for Research on Evaluation, Standards, and Student Testing at CRESST.

### **William Mehrens, Michigan State University**

Professor William Mehrens is professor emeritus of measurement and quantitative methods in the College of Education at Michigan State University. He serves on technical advisory committees in Michigan, Georgia, Arizona, Arkansas and Ohio. His research areas include educational testing, legal issues in high-stakes testing, teaching to tests, and performance assessment. He is past president of the National Council on Measurement in Education and the Association for Measurement and Evaluation in guidance, and served as vice president of Division D of the American Educational Research Association.

### **Edys Quellmalz, SRI International**

Dr. Quellmalz is Associate Director of the Center for Technology in Learning at SRI International. She conducts research, evaluation, and development projects related to the designs of technology-supported student assessments. She leads assessment projects funded by NSF and the U.S. Department of Education, involving the creation of network-based systems for evaluation and performance assessment resources and for the development of innovative, technology-supported student assessments. She is currently co-directing a project, funded by the National Science Foundation, entitled "Validities of standardized-based science inquiry assessments". She led the development of the NSF-funded project, "Performance Assessment Links in Science (PALS), an online standards-based collection of over 250 K-12 science performance assessments and accompanying professional development of assessment literacy.

### **Joseph Ryan, Arizona State University**

Professor Ryan is the Director of the Research Consulting Center at Arizona State University West in Phoenix. Professor Ryan has worked in the area of applied psychometrics for more than 20 years. He has extensive experience with the standard-setting process and procedures, including the Nedelsky, Angoff, and various other item-mapping [bookmarking] methodologies. Recently he contributed two chapters to the book, *Large Scale Assessment for All Students*, G. Tindel and T. Haladyna (Eds.), published by Lawrence Earlbaum Co. Professor Ryan has served as a technical advisory committee member in a dozen states and has worked as a consultant for most of the major testing companies.

### **Martha Thurlow, University of Minnesota**

Dr. Thurlow is the Director of the National Center for Educational Outcomes (NCEO). NCEO provides national leadership in designing and building educational assessments and accountability systems that appropriately monitor educational results for all students, including students with disabilities and students with limited English proficiency. NCEO works with states and federal agencies to identify important outcomes of education for students with disabilities, examines the participation of students in national and state assessments, including the use of accommodations and alternate assessments, evaluates national and state practices in reporting assessment information on students with disabilities, bridges general education, special education, and other systems as they work to increase accountability for results of education for all students. Dr. Thurlow serves on technical advisory committees in several states.

### **Joseph Willhoft, Tacoma Public Schools**

Dr. Willhoft is Executive Director of Research and Evaluation for Tacoma Public Schools. He has served on national technical advisory panels for Maryland, Nebraska, and Washington. Dr. Willhoft is past president of the Maryland Assessment Group, the Washington Educational Research Association, and the AERA Classroom Assessment Special Interest Group. He has been involved in Washington State's educational reform efforts for the past ten years, including service on committees for Washington's Commission on Student Learning, the state Accountability Task Force, and the Office of the Superintendent of Public Instruction. His special interests include the integration of assessment principles into school and district-based decision-making and effective graphical display statistical and assessment information.